

Statistical Research Report  
Institute of Mathematics  
University of Oslo

No. 4  
August 1965

A PRINCIPLE FOR CONDITIONING ON AN ANCILLARY STATISTIC

By

Else Sandved

## 1. Introduction and summary

In "Desisjonsteorien og Neyman-Pearson teoriens stilling idag", Sverdrup [3] considers in section II, C,D and E, conditional test methods, and he discusses the well known fact that the principle of unbiasedness in certain cases implies conditional tests. Sverdrup doubts, (illustrated by examples), that the principle of unbiasedness always is the real reason for choosing conditional tests. He suggests that the ancillary principle, introduced by Fisher [2], might be more basic. We exhibit, however, in section 3 of the present paper, an example which shows that one can get poor tests if one condition on a statistic which is ancillary in the commonly accepted sense of this term. We analyze the reason for this shortcoming, and we propose that the usual explanation of ancillary statistics should be supplemented, resulting in a definition in section 3 below.

## 2. Statistical decisions. Unbiasedness. Conditioning.

A statistical decision problem can usually be formulated in the following way:  $X$  is a stochastic variable with probability measure  $P$ . (In this paper the stochastic variables and the parameters may be multidimensional.) A priori  $P \in \mathcal{P}$ , where  $\mathcal{P}$  is a given class of probability measures. Based upon an observation of  $X$  we shall choose a subclass of  $\mathcal{P}$  which we believe contains  $P$ , or we shall give an estimate for  $P$ .

Suppose for example that the probability distribution of  $X$  is parametric and given by  $f(x;\omega)$ , where  $f$  is a known function of  $x$  when  $\omega$  is given.  $\omega$  is unknown, but a priori one knows that  $\omega \in \Omega$ , where  $\Omega$  is a given set. Let  $\theta$  be a known function of  $\omega$ , and let  $(H)$  be the range of  $\theta$  when  $\omega$  varies in  $\Omega$ .

In point estimation one gives an estimate of  $\theta(\omega)$ , in interval estimation one selects an interval and states that it contains  $\theta(\omega)$ . In hypothesis testing one specifies a subset  $\Omega^0$  of  $\Omega$  which is identified with the hypothesis, and by means of an observation of  $X$  the hypothesis is accepted, that is, one asserts  $\omega \in \Omega^0$ ; or it is rejected, that is, one asserts  $\omega$  not in  $\Omega^0$ .

An unbiased test is a test such that, by applying it, the minimum probability for rejecting the hypothesis if it is false, is not less than the maximum probability for rejecting it if it is true. The principle of unbiasedness sometimes implies conditional tests, given a statistic  $a(X)$ . That is,  $a(X)$  is regarded as given, and one constructs the test by considering the conditional distribution of  $X$ , given  $a(X)$ . Sverdrup [3] gives examples, showing that the principle of unbiasedness in some cases implies tests such that one condition according to statistical intuition (see the following examples 1 and 2), in other cases implies tests where one condition in disagreement with statistical intuition (example 3). He also shows that there are cases where it seems reasonable to use a conditional test, but where the principle of unbiasedness does not imply this conditional test (example 4).

Example 1.  $X_1$  and  $X_2$  are independently and Poisson distributed with parameters  $\lambda_1$  and  $\lambda_2$  respectively. The hypothesis is  $\lambda_2 \leq a \lambda_1$ , where  $a$  is given. The principle of unbiasedness implies conditional tests, given  $X_1 + X_2$ . Conditional distribution of  $X_2$ , given  $X_1 + X_2$ , is binomial and depends on  $\lambda_1$  and  $\lambda_2$  through  $\lambda_2 / \lambda_1$ . Moreover, there exists a uniformly most powerful conditional  $\alpha$ -level test.

Example 2. Given a double dichotomic frequency table and consider testing of independence, assuming multinomial distribution, or assuming two binomial distributions. The principle of unbiasedness implies conditional tests, given the marginals.

Example 3.  $X_1, \dots, X_n$  are independently and identically normally distributed with expectation  $\xi$  and variance  $\sigma^2$ . The hypothesis is  $\xi \leq k \sigma^2$  against  $\xi > k \sigma^2$ , where  $k$  is given.  $V = \sum X_i$  and  $Y = \sum X_i^2$  together are sufficient for  $(\xi, \sigma)$ . The principle of unbiasedness implies conditional tests, given  $Y$ .

Sverdrup [3] says\*):

---

\*) Translated by the present author.

# "Justification of conditioning"

Conditional testing has bothered the statisticians through many years. One has felt that one cannot arbitrarily condition the tests. One needs certain rules. A classical example is example 2 above with double dichotomy. Can one just assume that the marginals are given non-stochastical variables, regardless whether or not they are chosen in advance of the statistical experiment? Another classical example is regression analysis (see example 5 below). Should the independent variables be considered as "given" ("fixed") variables or as stochastical variables?

The problem has also been formulated as a problem of what the right sample space should be or what the "hypothetical repetitions" are.

If one is looking for a fixed rule, one can use unbiasedness. It entails conditional tests in the examples 1, 2 and 3 and in many other Poisson, multinomial and linear normal situations. It also underlies combinatorial testing in non-parametric situations. In those cases where statistical intuition seems to indicate conditional testing (examples 1 and 2), one is, however, not convinced that unbiasedness was the motive. Besides, one will perhaps feel that in some cases one is guided into wrong directions by the principle of unbiasedness (example 3).

The following example is illustrating:

Example 4 A quantity  $\xi$  is about to be measured to clear out if  $\xi = 0$  or  $\xi > 0$ . The result of the measurement is  $X$ . There are two instruments available for measuring  $\xi$ . According to instrument labeled 1  $X$  is normal  $(\xi, \sigma_1)$ , and according to instrument labeled 2  $X$  is normal  $(\xi, \sigma_2)$ , where  $\sigma_1$ , and  $\sigma_2$  are known. Let  $Y$  (=1 or 2) be the label of the instrument. One wants to seek out an institute which has one of the instruments, and it seems clear that one should assert  $\xi > 0$  if the instrument is labeled 1 and  $X > 1.64 \sigma_1$  or if the instrument is labeled 2 and  $X > 1.64 \sigma_2$  (5% level).

This is conditional testing given  $Y$ , and it seems obviously reasonable. But one can argue that one is in the "wrong" sample space. One should after all consider the sample space of  $(X, Y)$ . Let  $\Pr(Y=1) = p$ ,  $\Pr(Y=2) = 1-p$ . We assume  $p$  unknown. We have a nuisance parameter  $p$  in addition to the decision parameter  $\xi$ . By requiring unbiasedness one is lead to the just mentioned test.

But suppose now that one was informed that the institute, when buying the instrument, drew lots about the instruments with probability  $p = \frac{1}{2}$ . This should obviously be completely irrelevant. One can just look at the label on the instrument and ascertain which instrument the institute in fact had bought. The conditional test may still be reasonable.

But this result does not follow by applying the principle of unbiasedness. This principle is of help only if  $p$  is unknown. It is easily found that the most powerful test for the alternative  $\xi$  consists in rejecting the hypothesis if and only if

$$X > \frac{\xi}{2} + \frac{\sigma^2}{\xi} k, \text{ where } k \text{ is such that}$$

$$\frac{1}{2} \left[ 1 - G\left(\frac{\xi}{2\sigma_1} + \frac{\sigma_1}{\xi} k\right) \right] + \frac{1}{2} \left[ 1 - G\left(\frac{\xi}{2\sigma_2} + \frac{\sigma_2}{\xi} k\right) \right] = 0.05,$$

where  $G$  is the cumulative normal distribution. This test does depend on  $\xi$ , and there is no uniformly most powerful test. (See also Cox [1].)"

Consequently, in example 4, for the hypothesis  $\xi = 0$  against  $\xi = \xi_1$ , where  $\xi_1$  is given, the unconditional 5%-level test is more powerful than the conditional 5%-level test. Yet the conditional test is recommended, because one has a feeling of taking something irrelevant into account if one uses the unconditional test. It appears from what Cox [1] says in connection with this example that if  $p = \frac{1}{2}$  and  $\sigma_1$  is much greater than  $\sigma_2$ , then the unconditional 5%-level test for the hypothesis  $\xi = 0$  against  $\xi = \xi_1$ , considered as a conditional test, has level nearly 10% if  $\sigma = \sigma_1$ .

Cox [1] says: "Suppose that we know we have an observation from  $\Sigma_1$  (that is,  $\sigma = \sigma_1$ ). The unconditional test says that we can assign this a higher level of significance than we ordinarily do, because if we were to repeat the experiment, we might sample some quite different distribution. But this fact seems irrelevant to the interpretation of an observation which we know came from a distribution with variance  $\sigma_1^2$ . That is, our calculation of power, etc. should be made conditionally within the distribution known to have been sampled, i.e. if we are using tests of the conventional type, the  <sup>$\alpha$</sup> conventional test should be chosen."

We quote again from Sverdrup [3] \*):

"Example 5 Suppose that in the conditional distribution, given  $V_1, \dots, V_n$ , <sup>the variables,</sup>  $X_1, \dots, X_n$  are independently and normally distributed with variance  $\sigma^2$  and expectations  $\alpha + \beta V_i$ , respectively,  $i=1, \dots, n$ . We want to say something about  $\alpha, \beta, \sigma$ . If either (i) one does not know anything about the distribution of  $V_1, \dots, V_n$  or (ii) they are independently and normally  $(\nu, \tau)$  distributed, where  $\nu$  and  $\tau$  are unknown, then unbiasedness implies conditional testing, given  $V_1, \dots, V_n$ . If, however, (iv)  $V_1, \dots, V_n$  are independently and normally  $(0, 1)$  distributed, then conditional testing is not justified by unbiasedness. There are obviously other reasons for conditioning. The distribution of  $V_1, \dots, V_n$  does not depend on  $\alpha, \beta, \sigma$  in any of the situations, but still they are of importance for the testing. They have, as Fisher expresses it, about the same significance as the size of the sample and may therefore be considered as given."

### 3. Ancillary statistics.

The conception of "ancillary statistics" was introduced by R.A. Fisher [2]. If the probability model is parametric and one wants to say something about a parameter on the basis of an observation of the stochastic variable  $X$ , one may consider a statistic  $a(X)$  as given if  $a(X)$  is ancillary. There are different definitions of an ancillary statistic  $a(X)$ , but a common feature of the definitions is that the distribution of  $a(X)$  shall not depend on the parameter of interest (Fisher [2] and Cox [1]). Sverdrup [3] mentions condition (i) in the definition below. However:

---

\*) Translated by the present author.

there are situations where (i) alone implies poor tests (see example 6). Below, we explain the reason why (i) is not sufficient for conditioning, and we propose that also condition (ii) shall be satisfied in order to denote a statistic "ancillary" and consider it as given in a decision problem.

We assume the following:  $X$  is a stochastic variable defined on a sample space  $(\mathcal{X}, \mathcal{A})$  (that is a space  $\mathcal{X}$  where we have defined a  $\sigma$ -algebra  $\mathcal{A}$ ).  $X$  has probability measure  $P$ . A priori  $P \in \mathcal{P}$ , where  $\mathcal{P}$  is a given class of probability measures defined on  $(\mathcal{X}, \mathcal{A})$ . Let  $(H)$  be a set of indices and  $\{\mathcal{P}_\theta, \theta \in (H)\}$  be a family of non-empty subclasses of  $\mathcal{P}$  such that if  $P \in \mathcal{P}$ , then  $P$  belongs to one and only one  $\mathcal{P}_\theta$ . We want to make inferences on  $\theta$  on the basis of an observation of  $X$ .

For instance, if we want to test a hypothesis, we may let  $(H)$  consist of two elements,  $\theta_0$  and  $\theta_1$ , where  $\mathcal{P}_{\theta_0}$  is the class corresponding to the hypothesis, and  $\mathcal{P}_{\theta_1}$  is the class corresponding to the alternative. We then have to choose between  $\theta_0$  and  $\theta_1$ , i.e. to accept or to reject the hypothesis.

Let  $a(X)$  be  $\mathcal{A}$ -measurable.  $a(X)$  induces a sub- $\sigma$ -algebra  $\mathcal{A}^a$  of  $\mathcal{A}$ . Let  $P^a$  be the measure  $P$  restricted to  $\mathcal{A}^a$ , and let  $\mathcal{P}_\theta^a$  be the class of all  $P^a$  with  $P \in \mathcal{P}_\theta$ .

Let  $a(X)$  be a statistic such that

(i) The classes  $\mathcal{P}_\theta^a$ ,  $\theta \in (H)$ , are identical.

(ii) The class of conditional probability distributions of  $X$ , given  $a(X)$ ,  $\theta$  and  $P^a$ , is independent of  $P^a$ .

We then define  $a(X)$  to be an ancillary statistic for the decision problem, and we propose the following principle:

In the decision problem at hand, start with the conditional distribution of  $X$ , given  $a(X)$ .

REMARK. Later on in this paper, we shall make a comment on the problem arising when there are several ancillary statistics

The motivation for the principle is as follows: As a consequence of (i) the knowledge of  $a(X)$  will give us no information about  $\theta$ . So far we might use the observation of  $a(X)$  as the starting point and consider the conditional probability dis-

tribution of  $X$  given  $a(X)$  when we want to make inferences on  $\theta$ . But when  $a(X)$  is observed, we will in general be able to assign subclasses  $\mathcal{P}^*$  of  $\mathcal{P}$  where it is reasonable to believe that the unknown  $P$  is. Then the following problem arises: Should we choose a very extensive  $\mathcal{P}^*$  where we are almost sure that  $P$  is, or should we choose a less extensive  $\mathcal{P}^*$  where we are not that sure that  $P$  is? Whichever  $\mathcal{P}^*$  we might choose, there would in general be subclasses of  $\mathcal{P}^*$  where we would believe more strongly that  $P$  is, than in the rest of  $\mathcal{P}^*$ .

Thus the structure of this situation is entirely different from the a priori situation, where we had given a class  $\mathcal{P}$  which we knew contained  $P$ , but where we formally did not assign any subclasses of  $\mathcal{P}$  where we thought it more or less likely for  $P$  to be.

Example 6.  $X_1, \dots, X_n$  are independently and identically normally distributed with expectation  $\xi$  and variance  $\sigma^2$ .  $\xi$  and  $\sigma$  are unknown. The hypothesis is  $\xi \leq 0$  against  $\xi > 0$ . The condition (i) (but not (ii)) is satisfied for the statistic  $(|\bar{X}|, \sum (X_i - \bar{X})^2)$ , where  $\bar{X} = \frac{1}{n} \sum X_i$ . Hence this statistic gives us no indication whether the hypothesis is true or false. But if we test conditionally, given this statistic, assuming nothing more about  $\xi$  and  $\sigma$  than we did a priori, the uniformly most powerful conditional  $\alpha$ -level test is the following: reject with probability  $2\alpha$  if and only if  $\bar{X} > 0$ . However, after having observed  $|\bar{X}|$  and  $\sum (X_i - \bar{X})^2$ , we are no longer completely ignorant about  $\xi$  and  $\sigma$ . We can assign intervals around  $+\bar{X}$ , and  $-\bar{X}$ , and an interval around  $\frac{1}{n} \sum (X_i - \bar{X})^2$ , where we have reasons to believe that  $\xi$  and  $\sigma^2$ , respectively, are confined. If we take this information into consideration, we would surely use a test which is more reasonable than the test just mentioned.



The condition (ii) takes care of this difficulty. When  $a(X)$  is observed, we are no longer entirely ignorant about  $P^a$ . The information that  $a(X)$  gives us on  $P$  do we get only through  $P^a$ , we are completely ignorant about the rest. But since the conditional probability distribution of  $X$ , given  $a(X)$ ,  $P^a$  and  $\theta$ , does not depend on  $P^a$ , we have now, by observing  $a(X)$ , arrived at a situation where we, because of (i), do not know more about  $\theta$  than we did a priori, and where we, because of (ii), are unable to estimate some of the subclasses of the class of all conditional probability distributions of  $X$  given  $a(X)$ , as more likely than other subclasses.

Hence there is no reason why we should not start by considering  $a(X)$  as observed, and if we do this, we have in addition eliminated an irrelevant "part" of  $\mathcal{P}$ , namely the "part" which appears when  $P^a$  varies. Hence we may specify  $P^a$  and consider it as given, and we may then start the decision problem by assuming  $a(X)$  as given.

Both (i) and (ii) are satisfied for  $a = X_1 + X_2$  in example 1, for  $a = Y$  in example 4 and for  $a = (V_1, \dots, V_n)$  in example 5, and conditioning on  $a$  is in accordance with the usual procedure in these cases. In example 3, (i) is not satisfied for  $a = Y$ , and conditional tests, given  $Y$ , does not seem very convincing either.

Example 2 deserves a special comment. (i) is not in general satisfied for the set of those marginals which are stochastic variables. Suppose for example that  $X$  and  $Y$  are independent,  $X$  is binomial  $(m, p)$  and  $Y$  is binomial  $(n, q)$ , where  $0 < p, q < 1$ . Consider the hypothesis  $p < q$  against  $p > q$ .

If  $m=n$ , then (i) is satisfied for  $a = X+Y$ . As for (ii), suppose that the distribution of  $a$  is completely specified and  $p < q$ . Then  $p$  and  $q$  are known, and the conditional distribution of  $X$  given  $a$  does depend on  $p$  and  $q$  through  $\frac{p}{1-p} \frac{1-q}{q}$ . Hence (ii) is not satisfied. Yet it seems reasonable to condition on  $a$ . The reason for this is the following: Let  $a$  be given

$p < q$  and the distribution of  $a$  be roughly described, and consider the corresponding class of conditional distributions of  $X$  given  $a$ . For other roughly described distributions of  $a$  we get other classes of conditional distributions of  $X$  given  $a$ . These classes are involved in each other to a much higher degree than for instance the two classes of probability distributions of  $X$ , for  $p < p_0$  and  $p > p_0$ , respectively, where  $0 < p_0 < 1$ . Hence (ii) is not very far from being satisfied.

If  $m \neq n$ , then (i) is not satisfied for  $a = X+Y$ , but the two classes of distributions of  $a$  for  $p < q$  and  $p > q$ , respectively, are strongly involved in each other.

The principle implies that if  $a_1$  and  $a_2$  both are ancillary, and  $a_2$  is a function of  $a_1$  (i.e.  $\mathcal{A}^{a_2} \subset \mathcal{A}^{a_1}$ ), then one should condition on  $a_1$ . Because, if one conditions on  $a_2$ , i.e. considers conditional distribution of  $X$ , given  $a_2$ , then  $a_1$  is ancillary in this situation, and by applying the principle once more, one is lead to condition on  $a_1$ . Thus if there are two ancillary sub- $\sigma$ -algebras (i.e. sub- $\sigma$ -algebras induced by ancillary statistics), such that one of them is contained in the other, one should condition on that one which induces the finest (i.e. the most comprehensive) sub- $\sigma$ -algebra. Unfortunately, in general there exists no finest ancillary sub- $\sigma$ -algebra.

Example 7.  $X$  and  $Y$  are variables such that  $\Pr(X = x, Y = y) = p_{x,y}$ ;  $x=0,1$ ,  $y=0,1$ . The hypothesis is  $p_{0,0}=p_{1,1} = \frac{1}{2}$  against  $p_{0,1}=p_{1,0} = \frac{1}{2}$ .  $X$  and  $Y$  are individually ancillary, and there is no other ancillary statistic (except the ancillary statistic which induces the trivial sub- $\sigma$ -algebra).

Hence, if  $a_1$  and  $a_2$  are individually ancillary, then  $(a_1, a_2)$  is not in general ancillary, and we have to choose between them (and possible other ancillary statistics) by one or another principle.

An open question is how one can find the ancillary statistic. Of particular interest is to find in a given situation criteria of the non-existence of ancillary statistics.

REFERENCES

- [1] D.R. Cox, "Some problems connected with statistical inference," Ann.Math.Stat., Vol.29 (1958)
- [2] R.A. Fisher, "The logic of inductive inference," J.R.Stat. Soc., B, Vol.16 (1954)
- [3] Erling Sverdrup, "Desisjonsteorien og Neyman-Pearson teoriens stilling idag," Nordic Conferance on Mathematical Statistics, June 1965, Aarhus, Denmark, also to be published in "Review of the international statistical institute."

\*\*\* \*\*  
\*\*